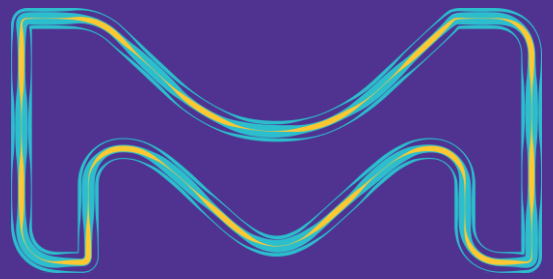
searching for novel chemical hit matter in large chemical spaces

NIH Virtual Workshop on Ultra-Large Chemistry Databases

Daniel Kuhn - Computational Chemistry
Merck Healthcare KGaA, Darmstadt Germany

Dec 1-3, 2020



Merck KGaA
Darmstadt, Germany

Thank you!

MERCK

Tim Knehans

Mireille Krier

Christina Schindler

Lukas Friedrich

Friedrich Rippmann

Vanita Sood

Cornelius Kohl

Gianna Pohl

Michael Krug

Jakub Gunera

Paul Czodrowski (TU Dortmund)

BioSolveIT

Franca Klingler
Christian Lemmen

IT-Consultants

Jan Fiedler

Christian Röder

Samo Turk

Andrew Dalke

GINKGO ANALYTICS

JKU JOHANNES KEPLER UNIVERSITÄT LINZ

Johannes Schimunek

Kristina Preuer

Günter Klambauer

Sepp Hochreiter

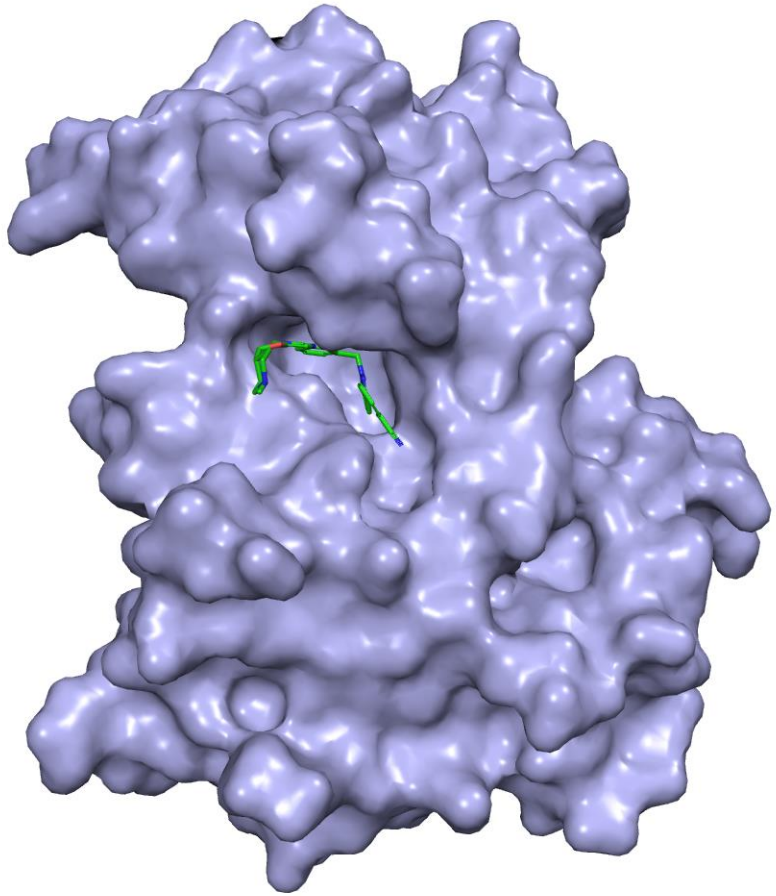
Compound optimization
Optimizing small molecule drugs – a multiparameter problem

Potency

Robust synthesis

Stability in vitro/vivo

IP – Freedom to operate



Physico-chemical properties

Solubility & permeability

Safety

Selectivity towards off-targets

Design, development and synthesis of drugs learned by medicinal and computational chemists and honed after years of training and practice.

Campbell et al. *Drug Discov. Today* **2018**, *23* (2), 219–234

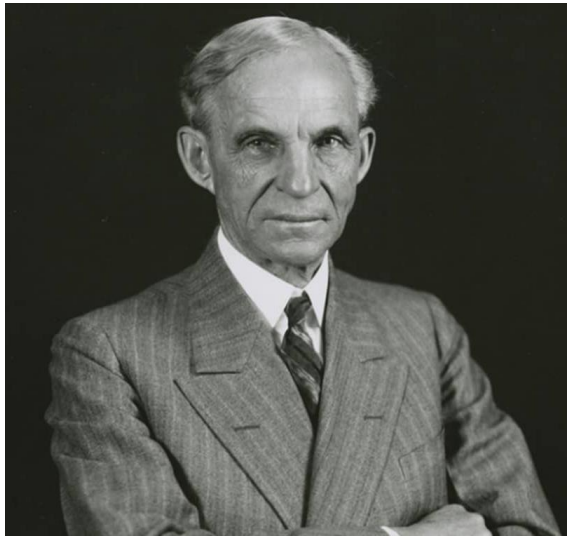
From art to process
Mission: Drive compound design by predictive modelling

Compound design and SAR analysis moving from art towards process

Designing better drugs faster: the patient is waiting

Mozart

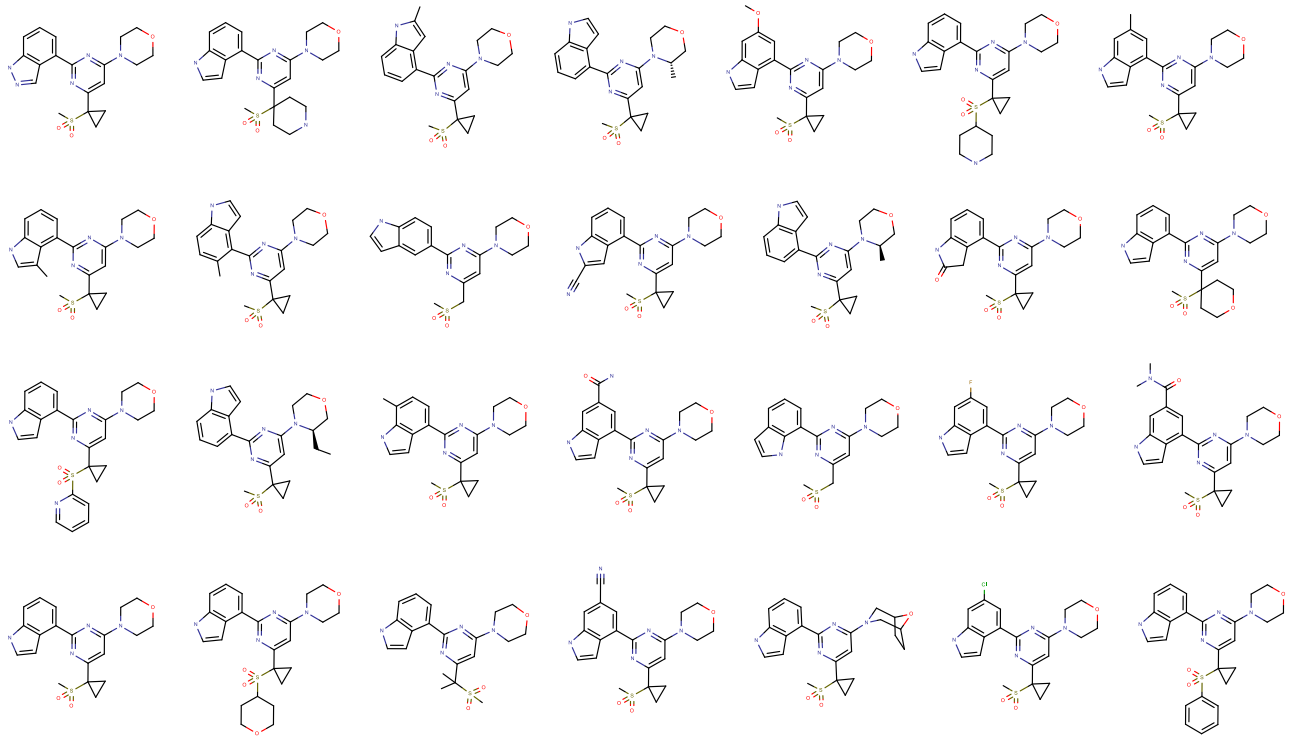
Ono

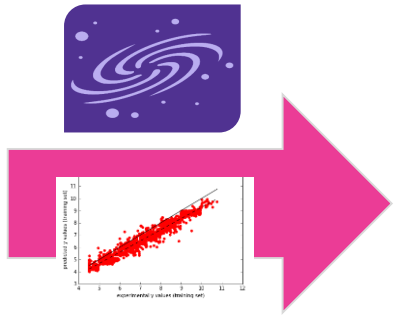


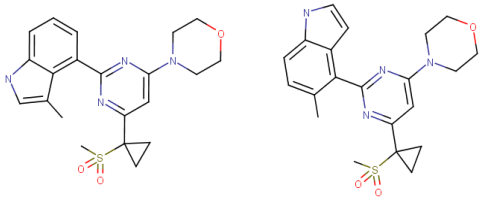
Ford

Which compounds to make next?
Challenge: Chemical space is huge – which starting point to pick?

Select most promising compounds out of huge chemical space, e.g.







Predicted with good activity, permeability, stability

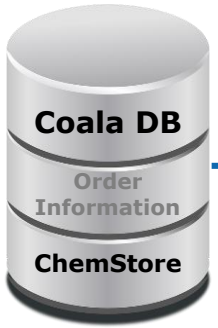
Inhibitors from Chembl database for protein target ATR (CHEMBL5024)

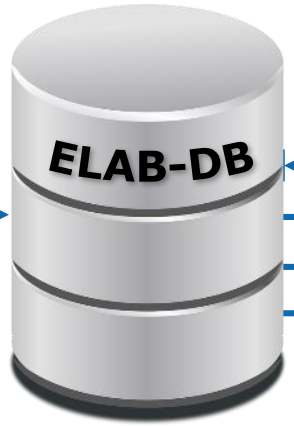
Foote et al., *J. Med. Chem.* **2013**, *56*, 2125–2138

> 20 years of Chemistry Electronic LAB Notebook

Elab as Merck-internal Knowledge Sharing Platform

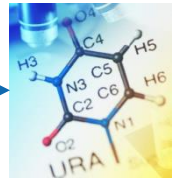
Reagent Stores





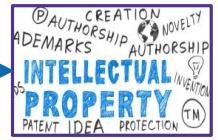
Order

Result



Analytic Systems
- **MassivR (site Da)**
- **ADMS (sites Da/Bi)**

Compound Registration (Biopharma)



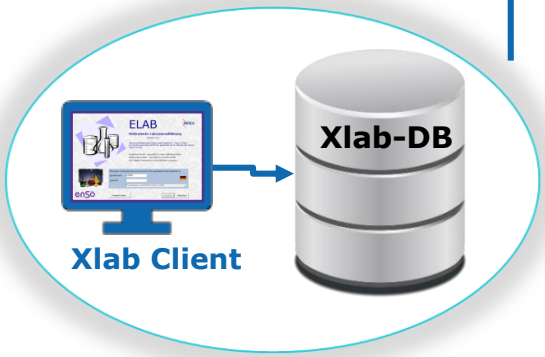
IP Record Store (Biopharma)

Patent Module

ELAB Client

- **> 1000 Users distributed over 7 sites**
- **> 1.000.000 experiments**
- **> 650.000 unique reactions**

Exchange with CROs



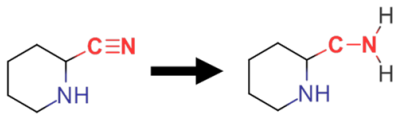
- **Documentation of chemical experiments**
- **Integrated workflows**
- **High quality data (contains also experiments that did not work)**
- **Creation and synthesis of libraries**
- **Knowledge sharing across business sectors**

- **Retrosynthesis**
- **Synthesis Design**
- **Virtual Library (MASSIV)**
- **Reaction Classification**

Analysis of ELAB (MERCK's ELN)

REACTION CLASSIFICATION

ELAB reaction



translation

- aliphatic carbon;not in ring
- tripple bond
- aliphatic nitrogen;not in ring

REACTS TO

- aliphatic carbon;not in ring
- single bond
- Nitrogen with three bonds;not in ring

SMARTS code

[C;!R]#[N;!R]>>[C;!R]-[NX3;!R]

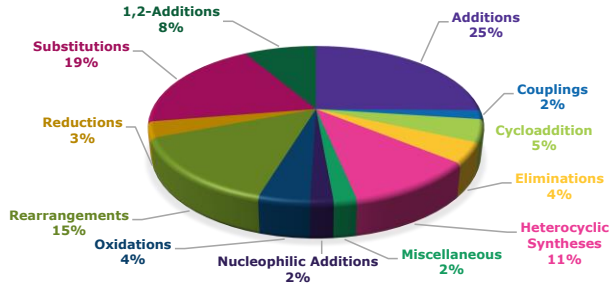
hashing

CLASSCODE

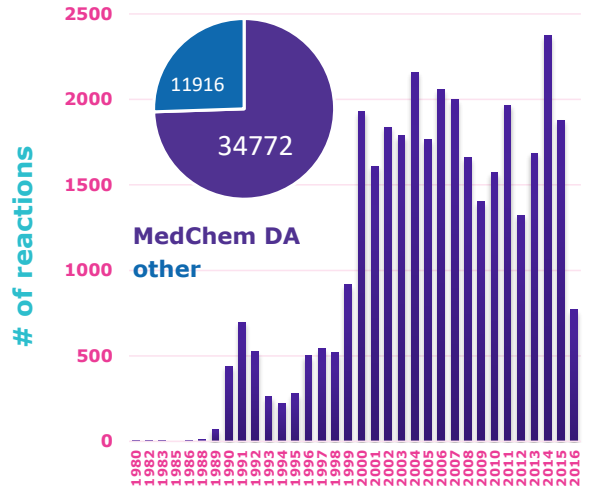
389072755771158

REACTION STATISTICS

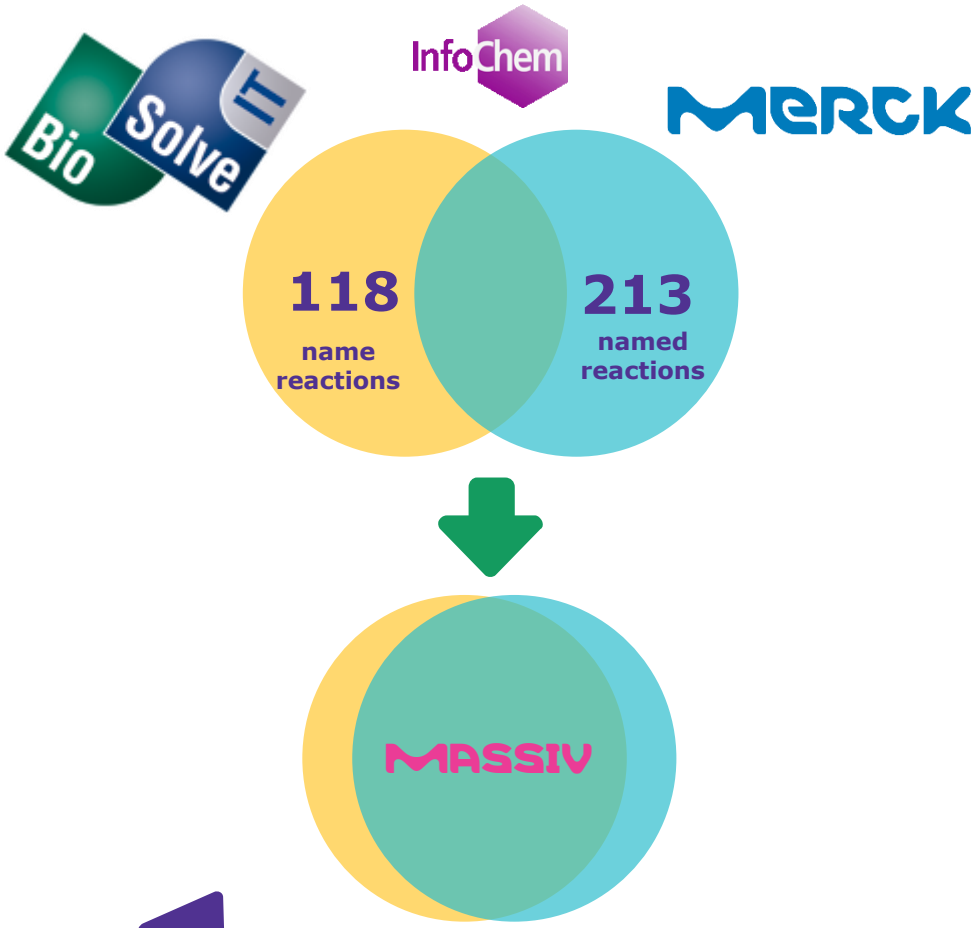
reactome analysis



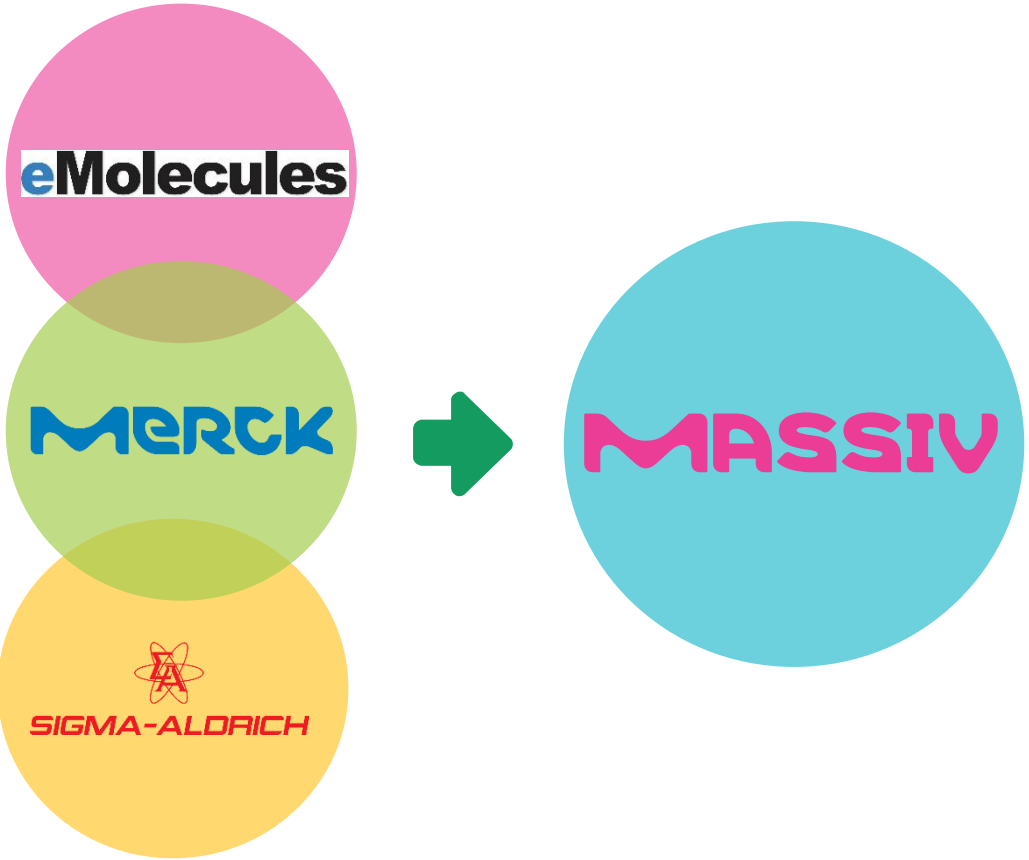
single-reaction analysis (261039242542204)



MASSIV overview of reaction space and building blocks



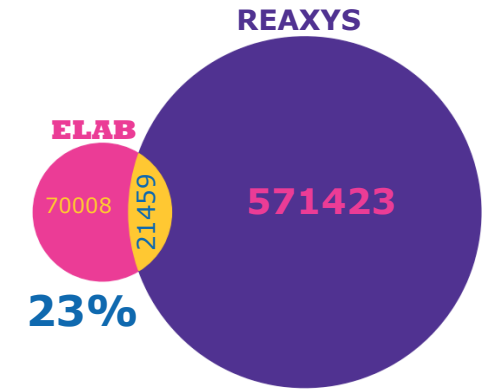
1 merging of reaction spaces



2 merging of building block libraries

Merck AcceSSible InVentory

CHEMICAL REACTIONS



10^4

in silico synthesis

BUILDING BLOCKS

10^6

MERCK

eMolecules

SIGMA-ALDRICH

BioSolveIT

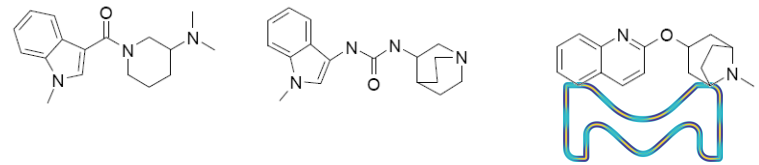
10^{20}

Virtual ChemSpace

LOOK-UP

novel chemical matter

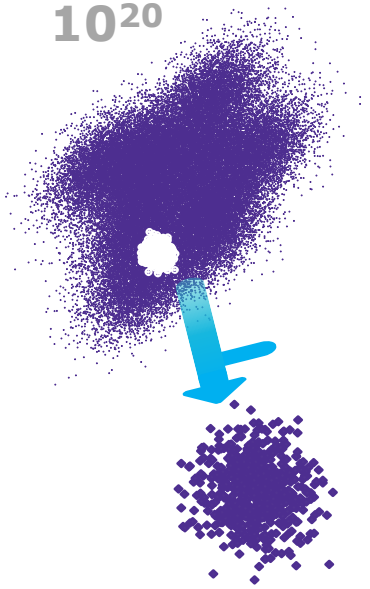
Tailored libraries

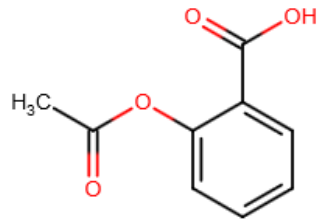


...

MASSIV space

10^{20}

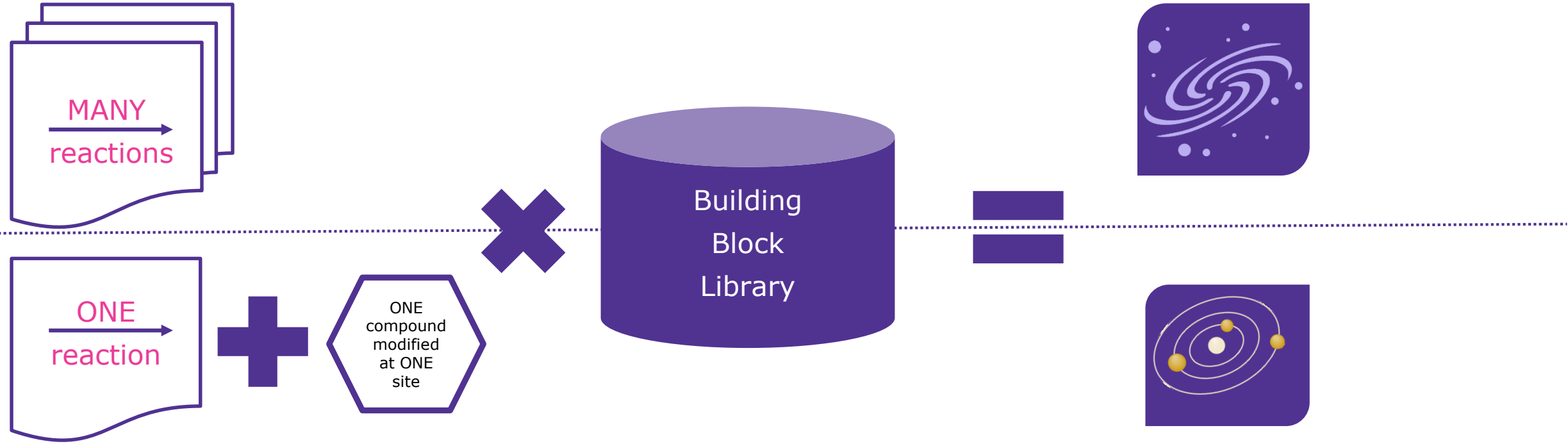




look-up space

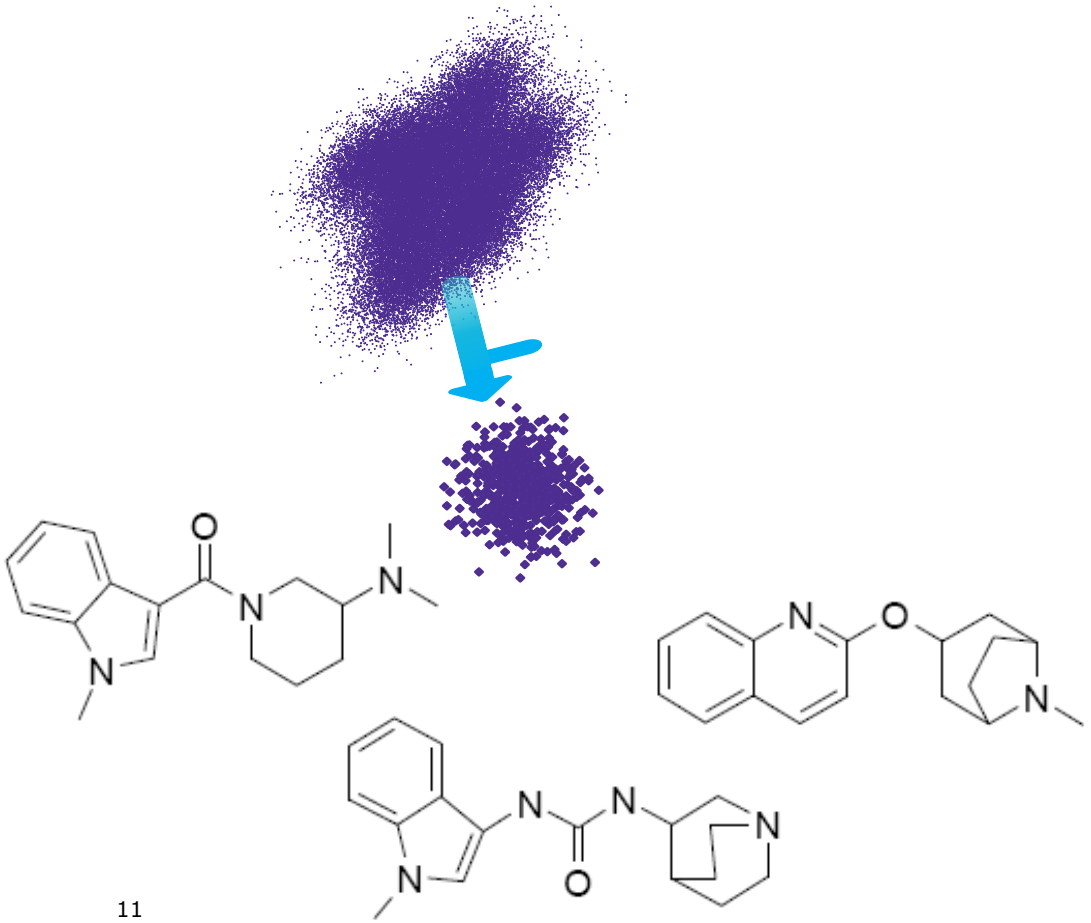
(10^{20} per reference)

MASSIV and miniMASSIV



Postfiltering important in hit selection

MASSIV space



FTrees

FPs

Docking

Shape

ADME

Clustering

Examples of MASSIV searches in ongoing Merck Projects

project	# Compounds analyzed	Synergies	Outcome
Receptor	~50,000	FEP, deep learning	
Enzyme	500,000	VS, deep learning, FEP	
Protein kinase	100,000	VS	Few actives found
Protein kinase	~1000	FEP, deep learning	Actives found
Enzyme	100,000	MedChem,VS	Actives found
Enzyme	32,000	VS, MedChem	
Protein kinase	100,000	VS, MedChem	Actives found
Enzyme	100,000	VS, MedChem	
Lipid kinase	100,000	VS frontloading	
Enzyme	100,000	VS frontloading	
Enzyme	~5000	VS, MedChem, FEP	50 actives found
Enzyme	~5000	VS, MedChem	
Enzyme	100000	VS, FEP	170 actives found
Enzyme	~50000	VS,MedChem	200 in synthesis

PoC - Synthesis impact achieved Actives found ongoing

Screen smarter not harder

Seit ich des Suchens müde ward,
Erlernte ich das Finden.
Seit mir ein Wind hielt Widerpart,
Segl’ ich mit allen Winden.

https://dribbble.com/naeememoradi

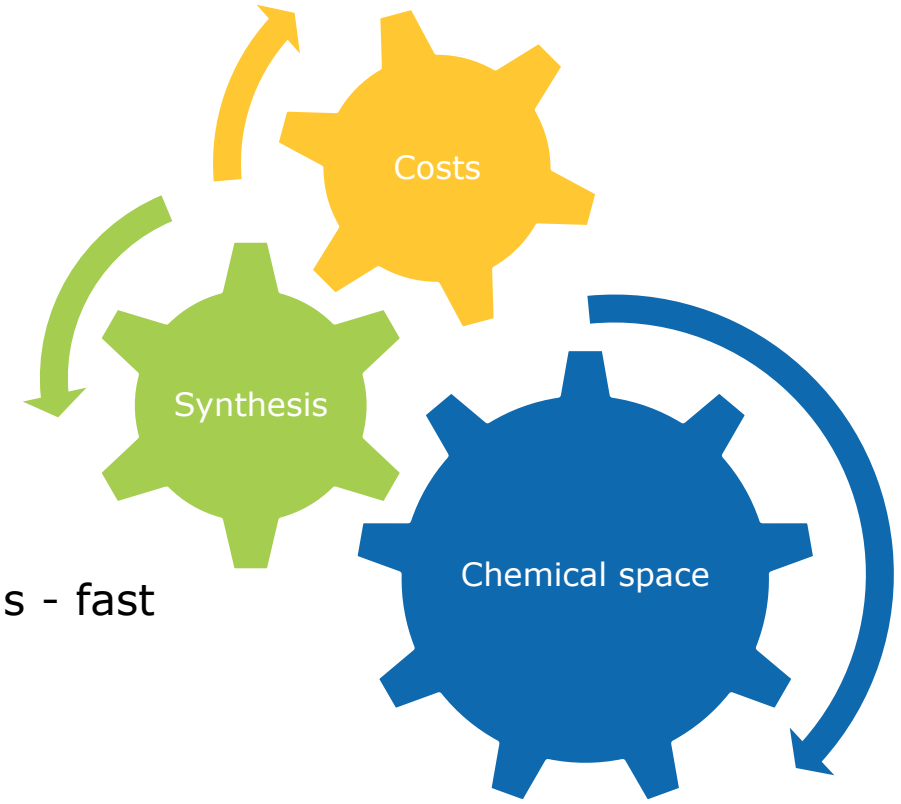
Since I grew tired of the chase
And search, I learned to find;
And since the wind blows in my face,
I sail with every wind.”

Friedrich Nietzsche - The Joyous Science

VS

Screen smarter not harder – lessons learnt

Ultra-large chemical spaces can provide interesting chemistry



Ingredients for quick take-up of virtual hits:

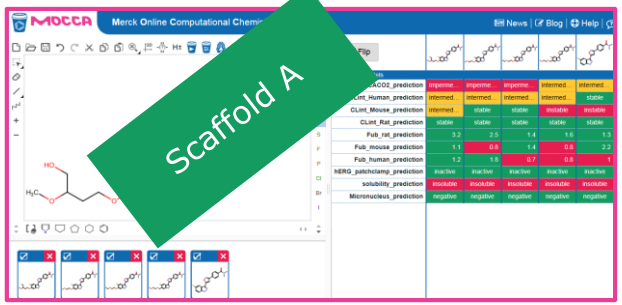
- Dedicated parallel chemistry resources to quickly follow-up on ideas - fast
- Out-sourcing to CROs – can be slow and expensive
- Search in dedicated make-on-demand chemical spaces from CROs (e.g. Enamine RealSpace) - fast

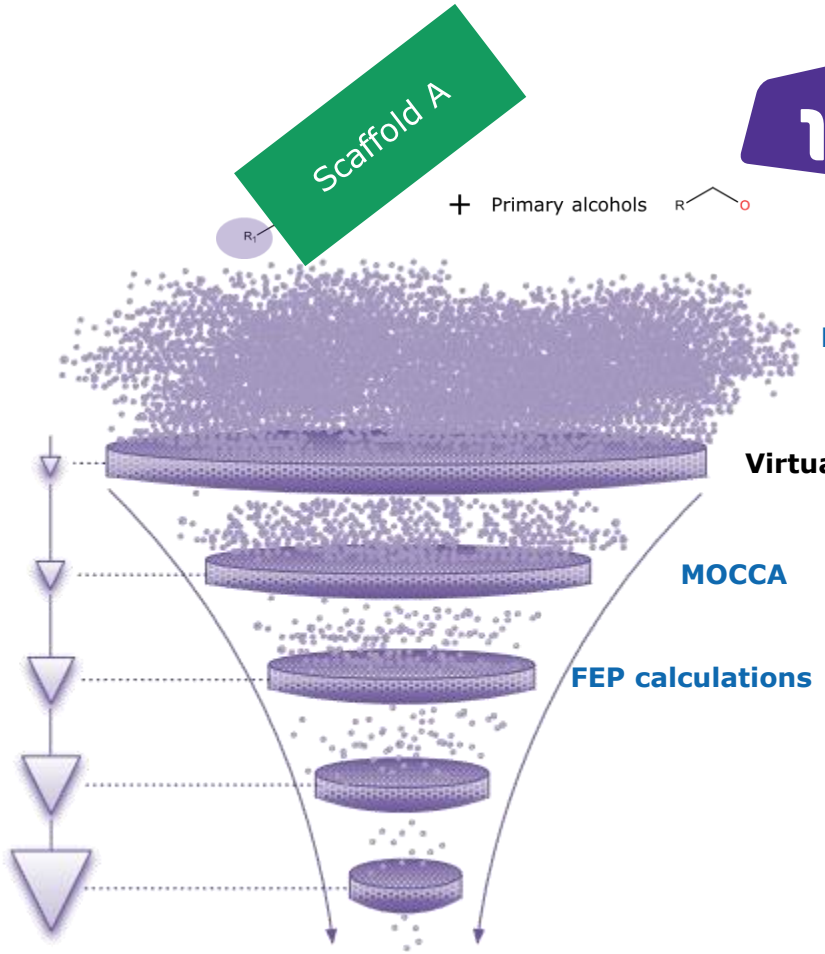
Combination is key for impact in compound optimization

2

MOCCA:

Application of predictive models



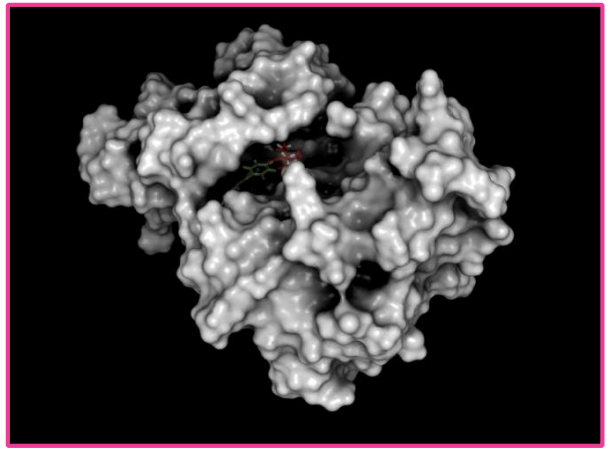


1

MASSIV: Enumeration of synthetically accessible chemical space

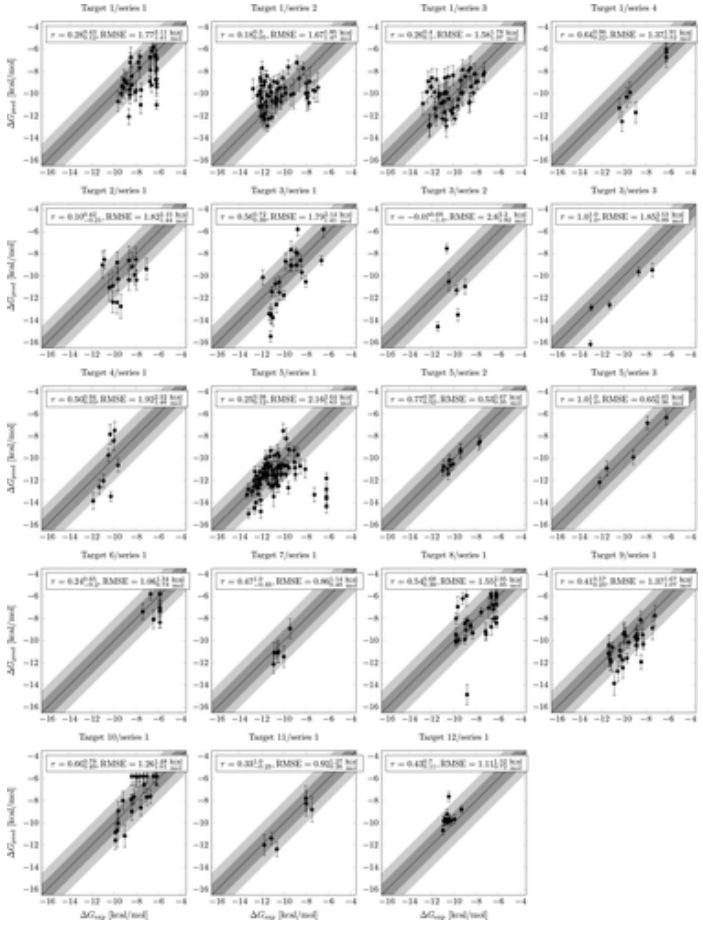
3

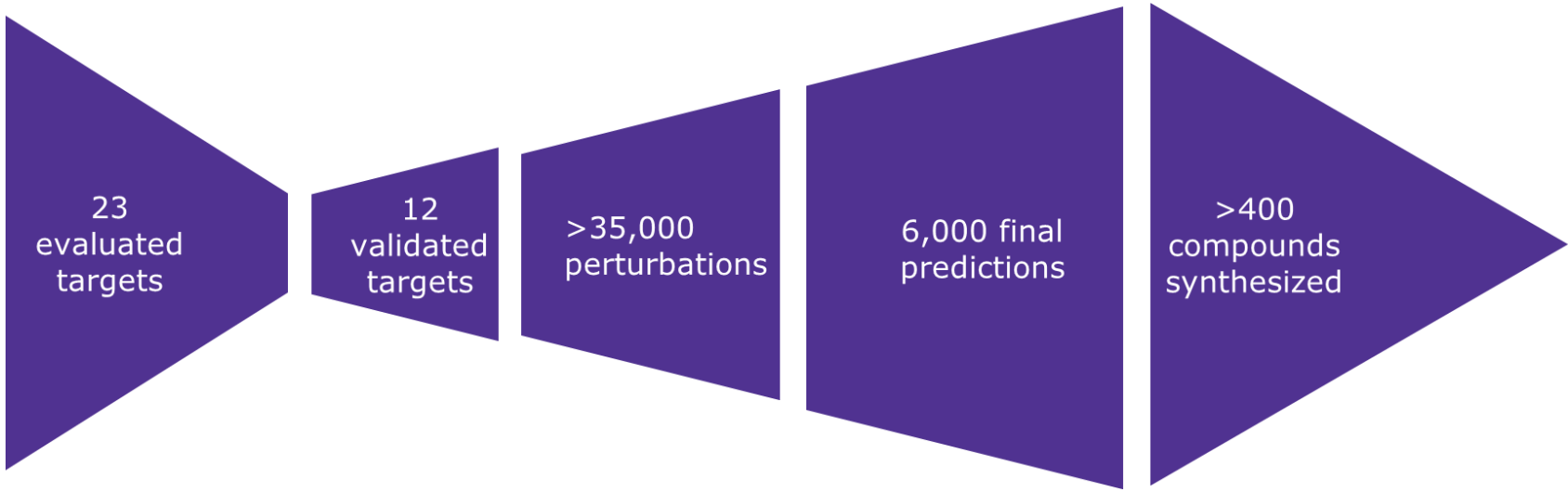
FEP: Binding constant prediction



Broad application across multiple targets and series

FEP+ in drug discovery at Merck





JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

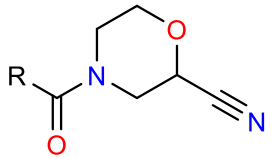
Large-Scale Assessment of Binding Free Energy Calculations in Active Drug Discovery Projects

Christina E. M. Schindler,* Hannah Baumann, Andreas Blum, Dietrich Böse, Hans-Peter Buchstaller, Lars Burgdorf, Daniel Cappel, Eugene Chekler, Paul Czodrowski, Dieter Dorsch, Merveille K. I. Eguida, Bruce Follows, Thomas Fuchß, Ulrich Grädler, Jakub Gunera, Theresa Johnson, Catherine Jorand Lebrun, Srinivasa Karra, Markus Klein, Tim Knehans, Lisa Koetzner, Mireille Krier, Matthias Leiendecker, Birgitta Leuthner, Liwei Li, Igor Mochalkin, Djordje Musil, Constantin Neagu, Friedrich Rippmann, Kai Schiemann, Robert Schulz, Thomas Steinbrecher, Eva-Maria Tanzer, Andrea Unzue Lopez, Ariele Viacava Follis, Ansgar Wegener, and Daniel Kuhn*

JCIM ASAP, 2020, https://doi.org/10.1021/acs.jcim.0c00900

Discovery of new chemical starting points with FEP+ML

From fragment to hit: Proof-of-principle for in silico optimization



SPR KDss = 300 µM

LE = 0.25

Enamine RealSpace
903 ideas

3D ROCS overlay
750 ideas

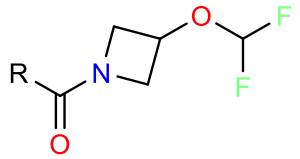
Docking + MMGBSA
400 ideas

ML model: CLint
250 ideas

FEP
8 ideas

Synthesis at Enamine
- 4 weeks
- < 100 EUR per compound

5 out of 8 molecules have IC50 < 100 µM

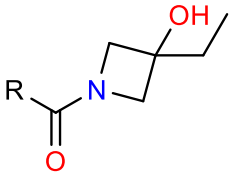


Top 1 in FEP

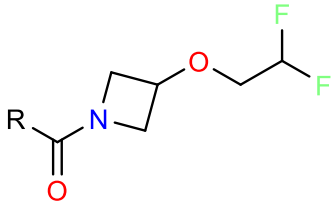
IC50 = 1.2 µM

ITC KD = 1 µM

LE = 0.41

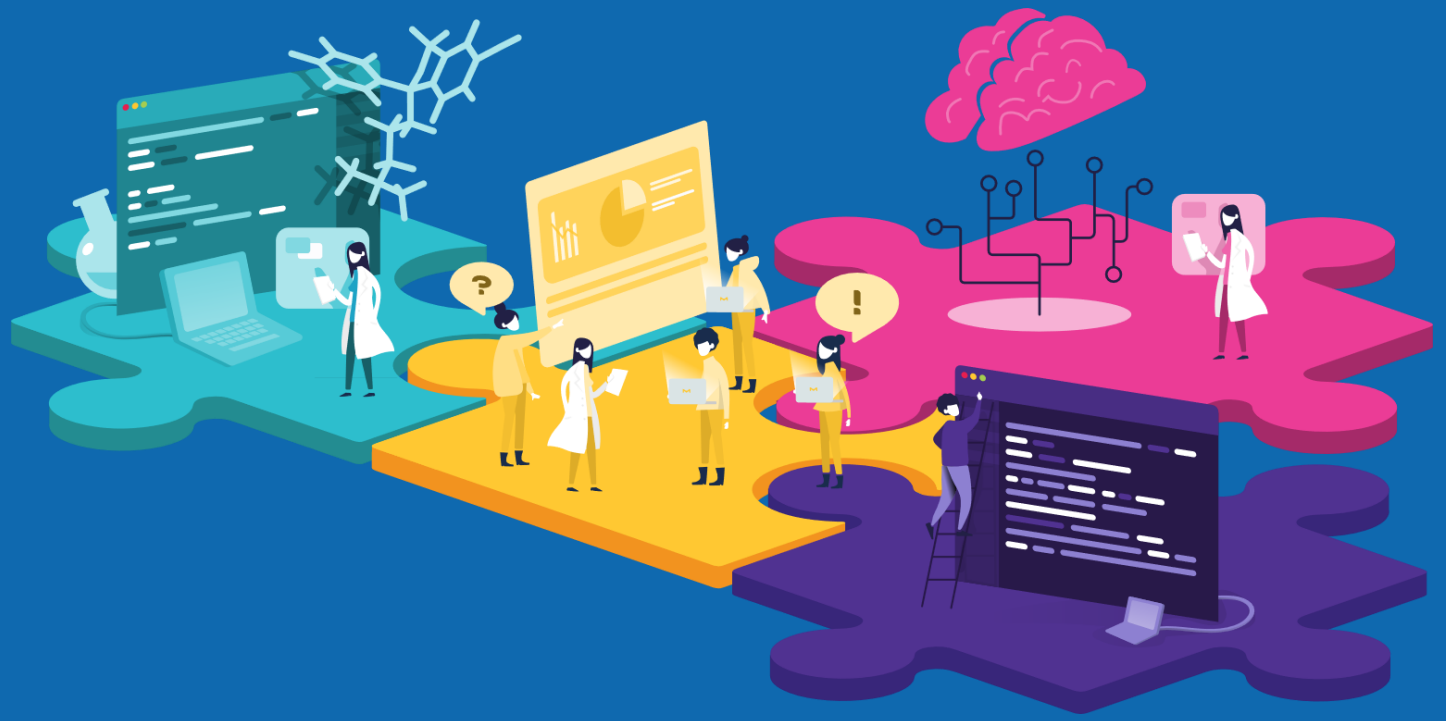


IC50 = 24 µM



IC50 = 47 µM

Christina Schindler



THANK YOU!

MERCK